AMENDMENTS TO THE SPECIFICATION:

Insert the following between lines 2 and 3 at page 3:

Figure 4 is a flow diagram illustrating a method for defining a normalized score according to another embodiment.

Figure 5 is a flow diagram illustrating a procedure for deriving a normalized score for a collection of profiles according to another embodiment.

Replace the numbered paragraphs 1-4 at page 5, lines 4-22 with the following:

1. The first parameter necessary to build the score is the number of thresholds used, denoted n. This count of thresholds includes all the (non-trivial) thresholds $r \in (0,1)$. Thus, if $r_0 \equiv 0$, the set of all thresholds in the score is indexed: r_k where $k = 0, 1, \ldots, n$. In addition, imposing $r_{n+1} = 1$ ensures that the highest tier's size is consistent with others. (Step 410 of Figure 4.)

2. Choose a mapping $r: \{1, 2, \ldots, n\} \rightarrow (0, 1)$ from the set of indices, as discussed below, to set all possible delivery-ratio thresholds. (Step 420.)

3. In the next step we need to run a set of trained profiles against the reference corpus. This run produces (via the delivery ratio mechanism used for each r_k for each profile) a set of vector space scores, x_k, $k = 1, 2, \ldots, n$ corresponding to the thresholds r_k for each profile in turn. (Step 430.)

4. Now, we are ready to score documents from the document source: Given an incoming document, we obtain a vector space score, x, of that document against a given profile, and we compare it to all the thresholds x_k. Let $l = 1, 2, \ldots, n-1$ be such that $x_l \leq x < x_{l+1}$, then we

assign a score $\sigma = 1$ to that document. For documents with $x < x_1$ we set the score $\sigma = 0$, and in case of $x \geq x_n$ the score is $\sigma = n$. This is the normalized score of the document against the given profile. Of course, the actual numerical value of σ can be defined in any number of ways, but the choice presented here is simple, and easy to interpret. (Step 440.)

Replace the numbered paragraphs 1-5 at page 11, lines 8-20 with the following:

1. Choose the number of uniform score levels to appropriately reflect the distribution of relevance over the source stream. (Set n.) (Step 510 of Figure 5.)

2. Decide on which mapping (for example, chose one of the two such mappings discussed above) is more appropriate for the given application. (Step 520.)

3. Select the threshold of relevance on the new scale for each tag. (Set $\theta = 3$. Optionally impose $r_3 = 0.9$; Instead of setting one of the r_k's to a given ratio, one could just fix the base of the exponent, which would in turn determine r_3.) (Step 530.)

4. Run the profile for the tag in retrieval mode against the reference collection in order to get the raw vector space scores for the delivery ratio thresholds comprising the score scale. (Step 540.)

5. For each new story, we can now determine its vector space score (step 550), compare it to the delivery ratio scores in the normalized score scale (step 560), and, depending on how many of those thresholds it passes, we can assign it a normalized (σ) score (step 570), and compare it to θ (step 580). This last comparison determines whether the news story receives the tag or not (step 590).

6. As an extra bonus, we end up with a rank-ordered list of tags. This can be used to limit the number of tags to a certain maximum (even if more tags manage to pass the

threshold of ÿ) of most relevant tags. It can also be used to assign several highest scoring tags to documents that did not score above ÿ for any tags, as long as some tags scored above 0.